

Chapter 2

Semantic Enhancement of Social Tagging Systems

Fabian Abel, Nicola Henze, Daniel Krause, and Matthias Kriesell

2.1 Introduction

Recent trends in the World Wide Web have shown an impressive growth of Web 2.0 systems, which are characterized by easy-to-use, interactive, and participatory usage scenarios. Users in Web 2.0 applications are more than ever active in the Web content lifecycle: They contribute with their opinion by annotating content (the so-called *tagging*); they add and annotate content (e.g., by using public, shared applications for their bookmarks, pictures, videos, etc.); they rate content or create content with sorts of online diaries (so-called *blogs*).

This new interactivity is possible by applications that are easy-to-use: users can use these applications right from the start; no remarkable training is required. Among the benefits for the users are of course, the gained interactivity and active participation and the possibility to profit from the commonly created knowledge: to search for content that has been annotated by other users with relevant tags, to explore new content by following *often used trails*, by digging into content that certain user groups assume relevant, and so forth. The collaboratively created, shared knowledge of a plethora of users provides interesting new means to detect, select, and recommend relevant knowledge items to Web users.

The Web 2.0 focusses especially on the usage dimension in the Web. Other dimensions such as the enhancement of the semantics dimension do improve accessibility and provide means to reason about Web content. Here, Web resources are embedded in a (machine understandable) context, where knowledge 04 bases

Fabian Abel, Nicola Henze, and Daniel Krause
IVS – Semantic Web Group, Leibniz University Hannover, Appelstr. 4, D-30167 Hannover, Germany, e-mail: {abel,henze,krause}@kbs.uni-hannover.de

Matthias Kriesell
Department of Mathematics, University of Hamburg, Bundesstraße 55 D-20146 Hamburg, Germany, e-mail: kriesell@math.uni-hamburg.de

V. Devedžić and D. Gašević (eds.), *Web 2.0 & Semantic Web*, Annals of Information Systems 6, DOI 10.1007/978-1-4419-1219-0_2,
© Springer Science+Business Media, LLC 2010

(so-called *ontologies*) provide pointers and references to both the content as well the context of Web resources and reveal important information.

Combining Web 2.0 ideas with semantic technologies gives benefits to both approaches. Semantic technologies supplement the *intercreativity* in Web 2.0 with expressive formats and languages to better employ and use created content and information. On the other hand, the Web 2.0 approach of easy participation provides possibilities to create valuable semantic metadata and, the Semantic Web still lacks sufficient (valuable) metadata.

In this article, we propose a novel way to combine semantic technologies and Web 2.0 paradigms. With the GroupMe! system [1], we have realized an appealing Web 2.0 application that enables users to easily construct groups of Web content that they consider interesting for some topic. GroupMe! users can group arbitrary Web resources such as videos, news feeds, images, etc. Within a GroupMe! group these resources are visualized according to their media type – e.g., videos can directly be played within a group, news feeds list their latest items, etc. – so that the content of groups is easy to grasp. GroupMe!’s tagging functionality allows users to annotate resources as well as groups. Hence, whenever resources are annotated, this is done in the context of a group.

The immediate benefit of the GroupMe! approach is that we are now able to see Web resources in a context, namely the group context: Web resources that were previously not related at all now have in common that they belong to some group which defines a common context. Together with tagging, we can even further specify this relation between the members of a group: The group’s tags are likely to be relevant for the members of the group, and vice versa. Our belief in this relevance can be specified by giving the relation between a member of a group and the group’s tags an appropriate weight. Thus, we capture the *semantics* of user interactions (creating a group, moving a Web resource into a group, resizing a Web resource, tagging it or the group, etc.) and produce – without additional overhead – valuable semantic metadata. Furthermore, groups of content provide us with a database of *hand-picked resources* for certain topics, which are specified by the group and its tags. Presumably, these resources are of high relevance for the topic – in comparison to search results lists – as a subject is screening the search results and decides which to add to the group, and which not.

In this article, we describe the GroupMe! system and investigate how to make use of this database of hand-picked resources and how to exploit the grouping structure on resources in order to improve the quality of ranking strategies in folksonomies. We benchmark our investigation against a popular ranking strategy in folksonomies, the FolkRank algorithm [2]. It turns out that the grouping structure significantly improves the quality of ranking.

The article is organized as follows: In the next section we introduce the GroupMe! system and architecture and present analysis describing the usage of the system. The captured semantic information is formally modeled in the GroupMe! folksonomy, which will be described and discussed in Sect. 2.3. Sect. 2.4 introduces and discusses ranking strategies in folksonomies, which are evaluated subsequently.

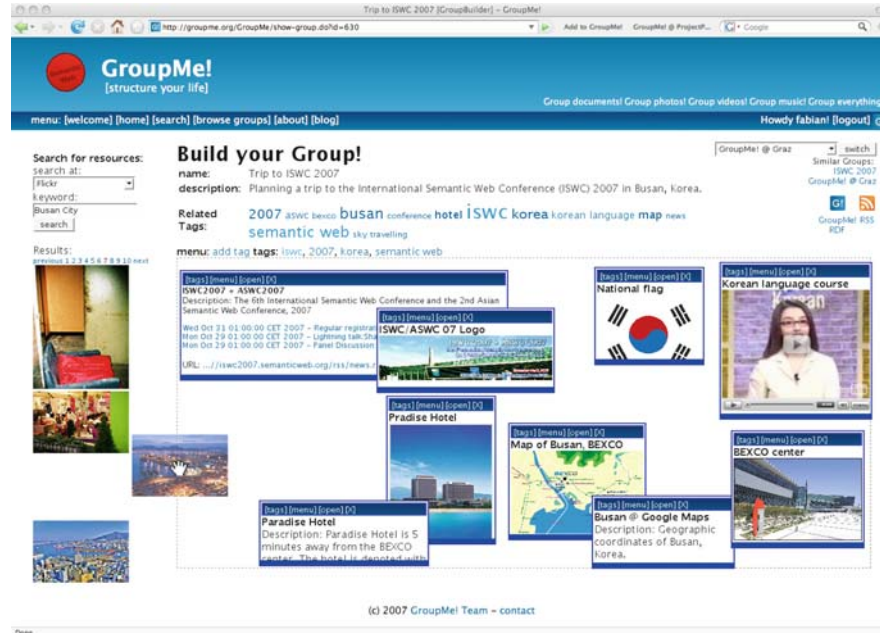


Fig. 2.1 Screenshot of GroupMe! application: A user drags a photo from the left-hand side Flickr search bar into the GroupMe! group on the right-hand side

In Sect. 2.5 we compare our approach to related work in the Semantic Web and Web 2.0. We conclude with an outlook on current and future work.

2.2 GroupMe! System

GroupMe!¹ is a new kind of resource sharing system. It extends the idea of social bookmarking systems with the ability to create groups of multimedia Web resources. Therefore, it provides an enjoyable interface, which enables the creation of groups via *drag & drop* operations. Resources within GroupMe! groups are visualized according to their media type so that users can grasp content without visiting each resource separately. GroupMe! groups form new sources of information as they bundle content, which is, according to the group creator, relevant for the topic of a group. GroupMe! groups are not only accessible for humans, but also for machines, because GroupMe! captures user interactions as RDF, i.e., whenever a user adds a resource to a group, annotates a resource/group, etc. GroupMe! produces RDF (see Sect. 2.2.1).

Figure 2.1 shows a screenshot of the GroupMe! system. It illustrates a scenario, in which someone utilizes the GroupMe! system in order to plan a trip to the ISWC

¹ <http://groupme.org>

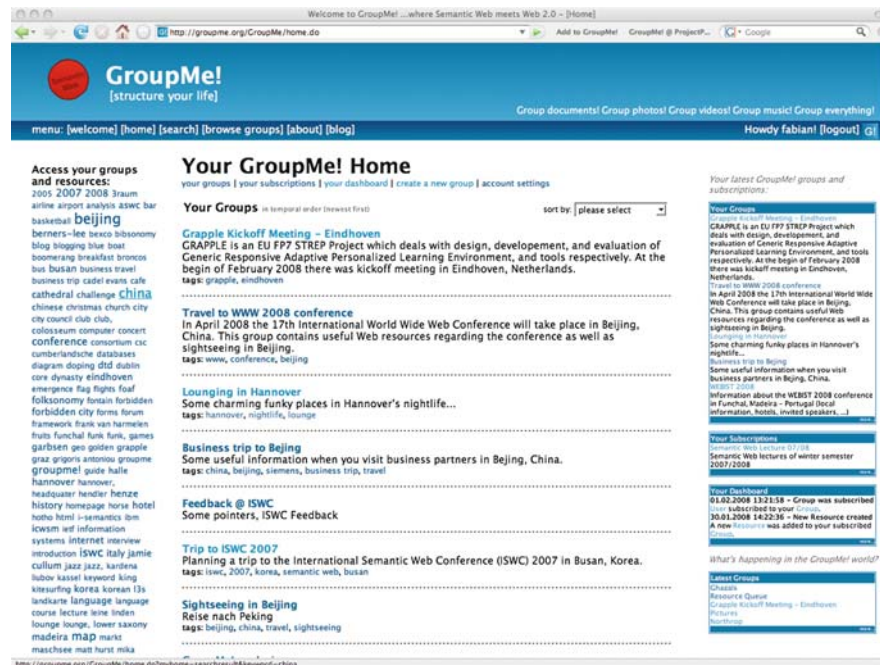


Fig. 2.2 Screenshot of the personal GroupMe! page. It lists the groups a user has created, groups she has subscribed to, events that recently occurred in groups of interest (*dashboard*), etc. Via a personal tag cloud (see left hand) the user can navigate to groups and resources she has annotated

2007 conference in Busan, Korea. Therefore, the user builds a GroupMe! group, which he names “Trip to ISWC 2007,” containing resources that are relevant for the trip. Building such a group is simple and can be done in two ways:

Browser Button. While browsing the web users can click on the GroupMe! browser button (*bookmarklet*) to add resources, they are interested in, to a group. When clicking the button users are directed to an input form where they can select the group(s) and specify tags they want to assign to the resource.

Group Builder. GroupMe! integrates different services such as Google or Flickr that enable users to discover and search for resources they may want to add to their groups. Figure 2.1 demonstrates how a user drags an image gathered from Flickr into his group. Drag & drop operations also allow to arrange resources within a group, i.e., to position and resize resources.

An important feature of the GroupMe! system is its visualization of groups. Resources are visualized according to their media type, e.g., pictures are displayed as thumbnails; videos and audio recordings can be played directly within the group, and RSS feeds are previewed by displaying recent headlines. Hence, content of GroupMe! groups can be grasped immediately. For example in Fig. 2.1, the Korean language video lecture can be watched instantly, the latest news about the conference is listed within the group, and photos of a hotel and the conference venue are

displayed. Altogether the arranged group in Fig. 2.1 appears like a *collage of information artifacts* about *ISWC 2007 trip* and gives an overview of the resources' content.

GroupMe! groups are interpreted as regular Web resources and can also be arranged within groups. This enables users to build hierarchies of Web resources and to make use of the information hiding principle – detailed information can be encapsulated into groups. Users who just want to get a rough overview about a topic do not need to visit those groups that contain detailed information.

GroupMe! groups are dynamic collections, which may change over time. Other users who also plan to attend the ISWC are enabled to subscribe to the group and will be notified at their *personal GroupMe! page* (see Fig. 2.2) whenever the group is modified, e.g., a new resource is added or removed, new tags have been assigned, etc. Users can also utilize their favored news reader to be up-to-date about changes within the group as each GroupMe! group provides an RSS feed. Thus, GroupMe! can be considered as a lightweight blogging tool where creation of blog entries is done via simple mouse operations instead of writing text. Information content is captured also by the group context, e.g., by adding the Web site “powerset.com” to a group “Promising Web 2.0 companies” the user denotes what he thinks about the corresponding company.

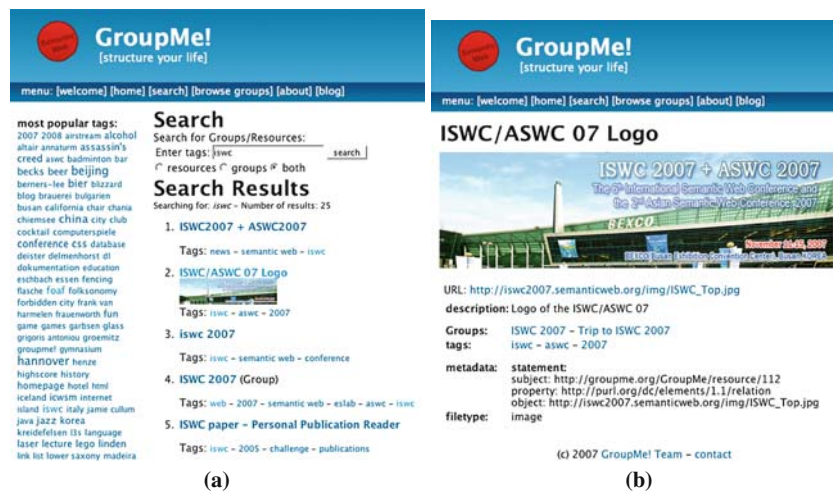


Fig. 2.3 (a) Search result list when searching for tag “iswc”, and (b) visualization of a single resource within the GroupMe! system

To ease future retrieval GroupMe! allows to tag both resources and groups. The personal GroupMe! page lists tags that a user has assigned to resources/-groups he is interested in: the *user tag cloud*. By clicking on a tag t within the user tag cloud he receives all resources/groups he has annotated with t . Tag clouds are furthermore computed and displayed for each GroupMe! group (see Fig. 2.1). Such group-specific tag clouds help users to get an overview about the topic of a

group. Another advantage of group-specific tag clouds is that they enable users to explore the GroupMe! corpus. Clicking on a tag t of a *group tag cloud* invokes a GroupMe! search operation, which results in a list of related resources and groups (see Fig. 2.3.a) – not only those resources that are directly tagged with t , as described in Sect. 2.4. Starting from a search result list, the user can navigate to other resources and groups (see Fig. 2.3.b). In general, all entities in GroupMe! – users, tags, resources, and groups – are clickable and resolvable, which results in an advanced browsing experience, e.g., each group points to similar groups (see top right in Fig. 2.1), or Resources refer to groups they are contained in.

Another important feature of GroupMe! is that content of groups is not only accessible and understandable for humans, but for machines, too. GroupMe! is there-with an RDF generator as it extracts RDF (meta)data about resources and captures each user interaction as RDF. RDF created in GroupMe! is made available to other Web applications and can be accessed via RSS and RDF feeds or RESTful [3] API. Hence, other applications can benefit from the feature of grouping and enriching resources with machine understandable semantics. The RDF generation functionality is described in more detail in the next section.

2.2.1 GroupMe! Architecture

GroupMe! is a modular Web application that adheres to the Model-View-Controller pattern. It is implemented using the J2EE application framework *Spring*.² Figure 2.4 illustrates the underlying architecture, which consists of four basic layers:

Aggregation. The aggregation layer provides functionality to search for resources a user wants to add into GroupMe! groups. Currently, GroupMe! supports Google, Flickr, and of course a GroupMe!-internal search, as well as adding resources by specifying their URL manually. *Content Extractors* allow us to process gathered resources in order to extract useful data and metadata, which are converted to RDF using well-known vocabularies.

Model. The core GroupMe! model is composed of four main concepts: *User*, *Tag*, *Group*, and *Resource*. These concepts constitute the base for the GroupMe! folksonomy (cf. Sect. 2.3). In addition, the model covers concepts concerning the users' arrangements of groups, etc. The *Data Access* layer cares about storing model objects. The actual data store backend is arbitrarily exchangeable. At the moment we are using a MySQL database.

Application logic. The logic layer provides various controllers for modifying the model, exporting RDF, etc. The internal GroupMe! search functionality, which is implemented according to the strategy pattern in order to switch between different search and ranking strategies, is made available via a RESTful API. It enables third parties to benefit from the improved search capabilities (cf. Sect. 2.4.3), and to retrieve RDF descriptions about resources – even such resources that were

² <http://springframework.org>

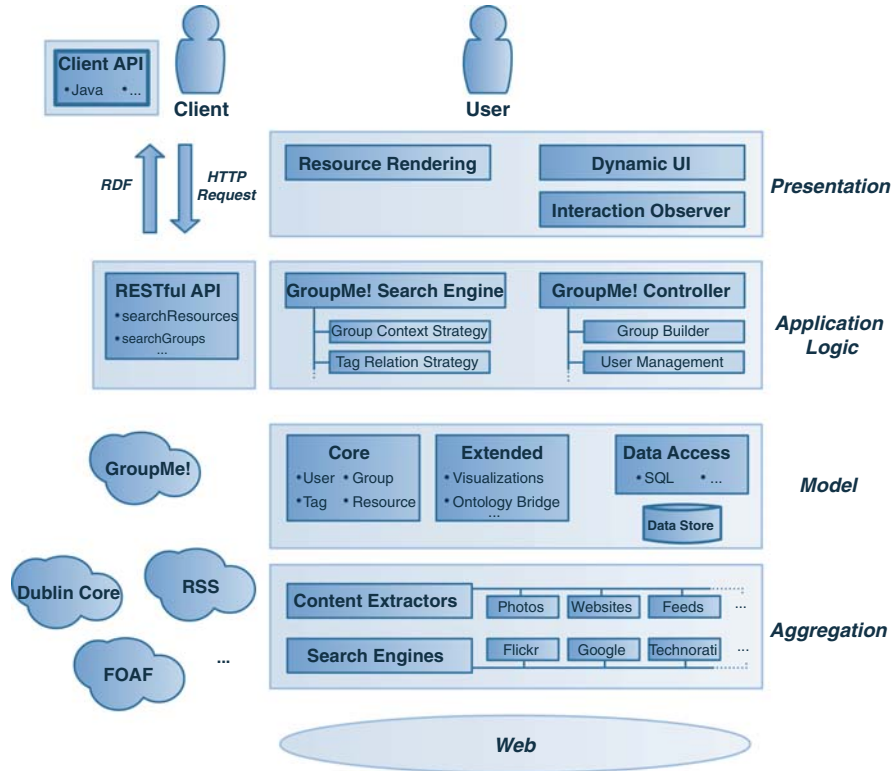


Fig. 2.4 Technical overview of the GroupMe! application

not equipped with RDF descriptions before they were integrated into GroupMe!. To simplify usage of exported RDF data, we further provide a lightweight Java *Client API*, which transforms RDF into GroupMe! model objects.

Presentation. The GUI of the GroupMe! application is based on AJAX principles. Therefore, we applied Ajax and JavaScript frameworks like `script.aculo.us`,³ `DWR`,⁴ or `Prototype`.⁵ Such frameworks already provide functionality to drag & drop elements, resize elements, etc. Visualization of groups and resources is highly modular and extensible. Switching between components that render a specific resource or type of resource can be done dynamically, e.g., visualization of group elements is adapted to their media type (see Fig. 2.1). In the future, visualizations of resources and groups should be adaptable by the users (see also Sect. 2.6).

When creating or modifying groups, each user interaction (e.g., moving and re-sizing resources) is monitored and immediately communicated to the responsible

³ <http://script.aculo.us>

⁴ <http://getahead.org/dwr>

⁵ <http://prototypejs.org>

GroupMe! controller so that e.g., the actual size or position of a resource within a group is stored in the database.

As mentioned in Sect. 2.2, GroupMe! is an RDF generator. RDF is generated with different methods.

1. Each user interaction (grouping and tagging) is captured as RDF using several vocabularies, e.g., *FOAF*⁶ and a GroupMe!-specific vocabulary⁷ that defines new GroupMe! concepts. External applications can therewith utilize information gained within the GroupMe! system like the information that two resources are grouped together or a certain tag was assigned to a resource within the context of a group.
2. Whenever a user adds a Web resource into a group, domain-dependent content extractors gather useful (meta) data so that resources can be enriched with semantically well defined descriptions. When, e.g., adding a Flickr photo into a group, a *Photo content extractor* translates Flickr-specific descriptions into RDF descriptions using *DCMI element set*.⁸ In the near future content extractors will be supported by frameworks such as *Aperture*,⁹ which facilitates extraction of data and metadata from different information systems and file formats.

Additionally we have implemented a (Meta) RDF search engine, which is currently added to the GroupMe! system in order to *query the Semantic Web* for existing RDF descriptions about resources, which are added to groups. Figure 2.5 illustrates the architecture of the search engine. It builds a wrapper around existing search engines such as *Sindice*,¹⁰ *Watson*,¹¹ or *Swoogle*,¹² and combines results of those engines. Mapping modules are utilized in order to adapt the result of such a combined RDF search result into a vocabulary that is understood by the requesting client, whereas the client's vocabulary capabilities are given as list of namespaces the client is aware of. At the moment we provide a module to translate from FOAF to Dublin Core vocabulary, and vice versa, e.g., we map (*#bob*, *foaf:knows*, *#mary*) to (*#bob*, *dc:relation*, *#mary*) and deliver both statements to the client.

To let other applications benefit from the semantics captured and gathered by the GroupMe! system, RDF descriptions of groups and resources are made available as feeds. An RDF description of group consists of, on one hand, RDF statements that describe the group itself and, on the other, statements, which describe the resources that are contained in the group. Figure 2.6 lists RDF produced by the current version of the GroupMe! system and represents parts of the RDF description of the group demonstrated in Fig. 2.1. The group is basically described with a name (*dc:title*), description (*dc:description*), and a list of contained resources

⁶ <http://xmlns.com/foaf/spec/>

⁷ <http://groupme.org/rdf/groupme.owl>

⁸ <http://dublincore.org/documents/dces/>

⁹ <http://aperture.sourceforge.net>

¹⁰ <http://sindice.com>

¹¹ <http://watson.kmi.open.ac.uk>

¹² <http://swoogle.umbc.edu>

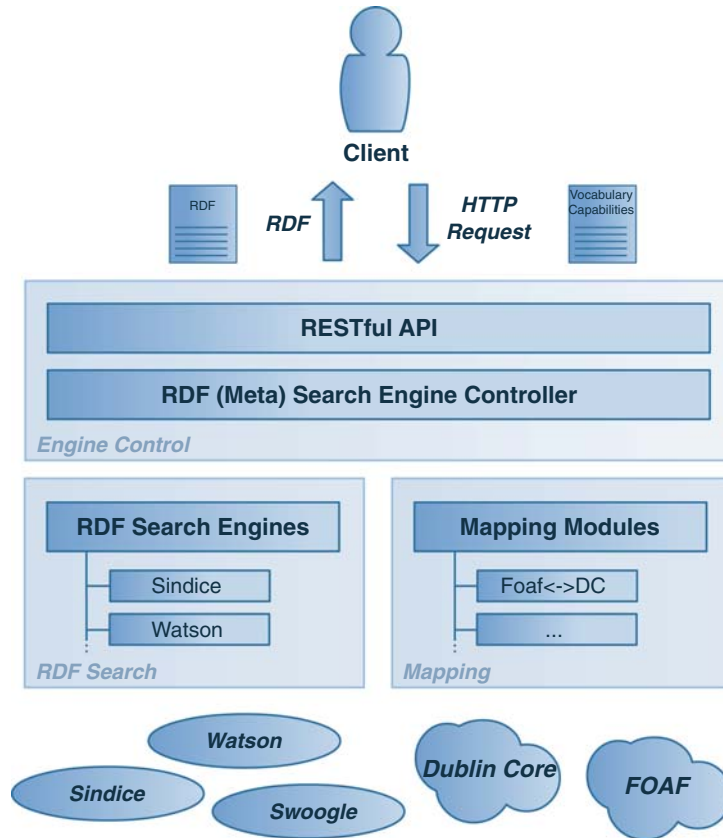


Fig. 2.5 Technical overview of the RDF (Meta) Search Engine

(`groupme:resource`). The extract of the group's RDF description specifies two resources, an image (`foaf:Image`), which represents a photo of a hotel in Busan, and a Web site (`foaf:Document`), which represents a certain Google map¹³. Via the concept *TagAssignment* it is stated which tags are assigned to groups/resources, e.g., the tag with id 90 was assigned to the photo within the *ISWC group* by user 3 on November 10, 2007. The photo also provides a simplified version of the formal tag assignment description by utilizing the attribute `dc:subject`. The Google map resource is even equipped with longitude (`wsg84:long`) and latitude information (`wsg84:lat`), whereby the location of Busan is specified precisely.

The RDF description displayed in Fig. 2.6 illustrates the main characteristics of the GroupMe! approach. By grouping resources, which were (possibly) not related

¹³ <http://maps.google.com>

with each other beforehand, they are set into a well-defined context, which enables applications to deduce additional knowledge. For example, as the photo and the Google map are thematically related to each other, the metadata that specify the geographic coordinates of the map may also be applicable to the photo of the hotel.

```

...
<groupme:Group rdf:about="http://groupme.org/GroupMe/group/630">
  <dc:title rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    Trip to ISWC 2007
  </dc:title>
  <dc:description >
    Planning a trip to the International Semantic Web Conference
    (ISWC) 2007 in Busan, Korea.
  </dc:description>
  <groupme:resource rdf:resource="http://groupme.org/GroupMe/resource/634"/>
  <groupme:resource rdf:resource="http://groupme.org/GroupMe/resource/653"/>
  <dc:subject rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    Semantic Web
  </dc:subject>
  ...
</groupme:Group>

<foaf:Image rdf:about="http://groupme.org/GroupMe/resource/653">
  <owl:sameAs rdf:resource="http://iswc2007.semanticweb.org/paradise.jpg"/>
  <dc:description rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    Paradise hotel in Busan
  </dc:description>
  <dc:title rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    Paradise Hotel
  </dc:title>
  <groupme:tas>
    <groupme:TagAssignment>
      <dc:created rdf:datatype="http://www.w3.org/2001/XMLSchema#date">
        2007-10-30
      </dc:created>
      <groupme:user rdf:resource="http://groupme.org/GroupMe/user/3"/>
      <groupme:tag tag="http://groupme.org/GroupMe/tag/90">
      <groupme:resource rdf:resource="http://groupme.org/GroupMe/
        resource/653"/>
      <groupme:group rdf:resource="http://groupme.org/GroupMe/group/630"/>
    </groupme:TagAssignment>
  </groupme:tas>
  <dc:subject rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    ISWC
  </dc:subject>
  ...
</foaf:Image>

<foaf:Document rdf:about="http://groupme.org/GroupMe/resource/634">
  <owl:sameAs rdf:resource="http://maps.google.de/maps?amp;q=Busan+Korea"/>
  <dc:description rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    Geographic coordinates of Busan, Korea.
  </dc:description>
  <dc:title rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    Busan @ Google Maps
  </dc:title>
  <wsg84:lat rdf:datatype="http://www.w3.org/2001/XMLSchema#double">
    52.393442
  </wsg84:lat>
  <wsg84:long rdf:datatype="http://www.w3.org/2001/XMLSchema#double">
    9.8061
  </wsg84:long>
  ...
</foaf:Document>
...

```

Fig. 2.6 Extract of RDF description of the group displayed in Fig. 2.1

An application that searches for photos according to a location specified via geographic coordinates is now able to retrieve photos by locations even if these photos are not directly annotated with geographic coordinates.

2.2.2 Evaluation of the GroupMe! System

In the previous section, we introduced the GroupMe! system and the concept of building groups of resources. To outline the benefit of our system, we evaluated how users interact with the GroupMe! system; in particular, we focused on usage and tagging characteristics, and investigated the effects of the structure given by the groups to search and retrieve resources. The data underlying the analysis was collected during the first six months after the system's launch on July 14, 2007. During the observed period, GroupMe! had a total of 1351 resources of which 1078 were normal resources and 273 (20.2%) were groups. Altogether, 1758 tag assignments were monitored, with 1.3 tags per resource in average. The overall evolution of resources and groups is given in Figure 2.7.

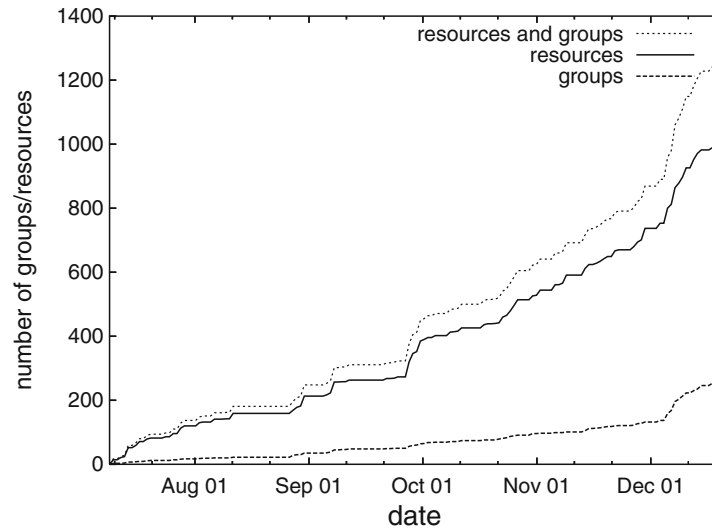


Fig. 2.7 Evolution of number of resources/groups

According to the *tagging system design taxonomy* proposed in [4], GroupMe! is a free-for-all tagging system, which allows users to annotate multimedia content for future retrieval. Hence, GroupMe! allows for broad folksonomies as every user is allowed to tag every resource or group without any restriction. Tagging a resource r is done when users are situated in the view of a certain group g . Thereby, users are only able to see those tags that have been assigned to r within the context of

the group g (same holds for group g). Explicit tag suggestions are not provided by the GroupMe! system. However, the tag cloud of a group and the resource's visualization, which is adapted to the media type of the resource, help the users to reflect on appropriate tags for the resource.

Interestingly, groups were tagged more intensively than ordinary resources: In average, 1.98 tags were assigned to groups, whereas only 1.13 tags were attached to other resources. Thus, groups were tagged 1.75 times more often than traditional resources. This effect was present over time, as depicted in Figure 2.8. Furthermore, at the end of the observed period only 36.98% of the groups were not annotated with any tag in contrast to 52.41% of the resources. These initial observations give support for the hypothesis that users adopt the group idea to organize Web resources, and that they also invest time in the group construction process.

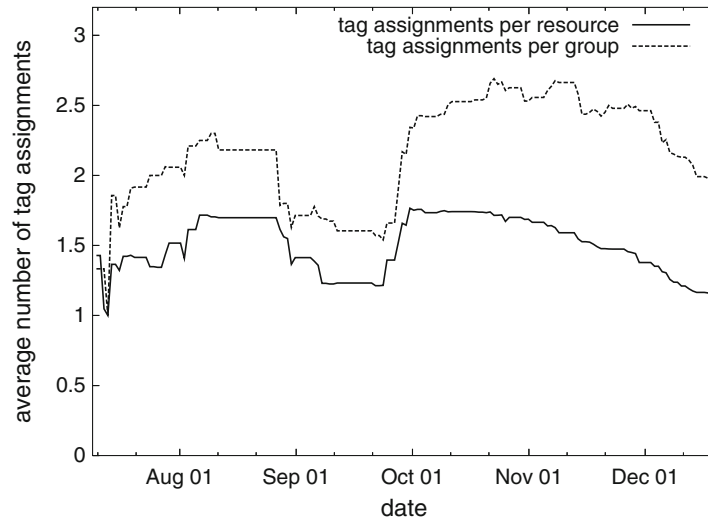


Fig. 2.8 Average number of tags assigned to resources/groups

A typical group in GroupMe! consists of 2–8 resources. That we do not observe groups with significantly more members can be explained from the user interface, which gives the users a canvas to place and arrange the Web resources. As the size of this canvas is limited, the on-screen display of the group becomes impractical with too many Web resources. Users collect resources with different media types in their group, as can be seen in Table 2.1. Most popular among the media types are images, followed by videos and RSS feeds. Web sites, academic papers, presentation slides, etc. are denoted as *other Web resources* and are not mentioned separately, because to users they appear as simple bookmarks, i.e., their visualization is not yet adapted to their media type particularly. The possibility to include groups into a group was only seldomly used; we explain this by the small number of available groups during the observation period.

Table 2.1 Percentage of resources' media types that are part of GroupMe! groups

Type of resource	AVG occurrences (%)
Images	39.77
Videos	5.66
Rss feeds	3.0
Groups	3.33
Other Web resources	48.22

2.2.2.1 Results

The evaluation of the GroupMe! system shows that users appreciate the grouping facility to organize Web resources they are interested in. Furthermore, we have shown that users benefit from the media independence of GroupMe!, as a rich mixture of media types is used in the GroupMe! system. Groups can be seen as *hand-picked* collections of Web content for a certain topic or domain. As such, they are also valuable results to perform search queries, which we investigate in the following sections.

2.3 GroupMe! Folksonomy

To develop FolkRank-based group-aware ranking strategies we have to embed the group context introduced by the GroupMe! approach into the formal folksonomy model. The term *folksonomy*, introduced by Thomas Vander Wal in 2004 [5], defines a taxonomy, which evolves over time when users (the *folks*) annotate resources with freely chosen keywords. Folksonomies can be divided into *broad* folksonomies, which allow different users to assign the same tag to the same resource, and *narrow* folksonomies, in which the same tag can be assigned to a resource only once [6]. Formal models of a folksonomy are, e.g., presented in [7] or [8]. They are based on bindings between users, tags, and resources. According to [9] a folksonomy is defined as follows:

Definition 1. A *folksonomy* is a quadruple $\mathbb{F} := (U, T, R, Y)$, where:

- U, T, R are finite sets of instances of *users*, *tags*, and *resources*, respectively.
- Y defines a relation, the *tag assignment*, between these sets, that is, $Y \subseteq U \times T \times R$.

In [10], tag assignments are furthermore attributed with a timestamp and Hotho et al. also embed relations between tags into the formal folksonomy model [9]. To simplify the formalization we do not include these features. GroupMe! introduces groups as a new dimension in folksonomies.

Definition 2. A *group* is a finite set of resources.

A group is a resource as well. Groups can be tagged or arranged in groups, which effects hierarchies among resources. In general, tagging of resources within the GroupMe! system is done in context of a group. Figure 2.9 presents a basic GroupMe! tagging scenario, in which users u_1 and u_2 have grouped resources r_{1-3} into g_1 and g_2 , and have tagged both resources and groups with keywords t_{1-3} . The tag assignment $tas_2(u_1, t_2, r_2, g_1)$ in Fig. 2.9 describes that user u_1 has annotated resource r_2 in context of group g_1 with tag t_2 . The group context of tags helps to detect ambiguous tags. For example, r_2 has also been tagged with t_2 in context of group g_2 , which indicates that the meaning of t_2 is probably the same in both groups.

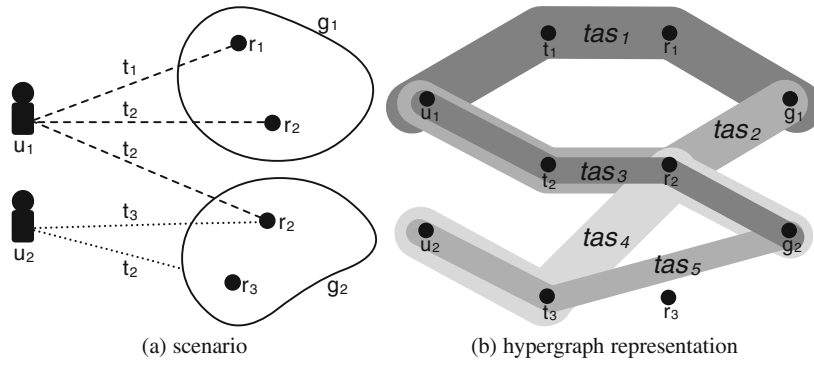


Fig. 2.9 (a) Scenario in which two users assign tags to resources in context of different groups and (b) the corresponding representation as a hypergraph

Assume that there is a group, which does not contain any of the resources of g_1 , and t_2 would be the only tag that occurs in both groups, then the meaning of t_2 is possibly ambiguous. If users assign tags to a group, which is itself not contained in a group, then the group context information is not available ($\rightarrow (u_2, t_2, g_2, \varepsilon)$) and within the hypergraph representation the tag assignment can be interpreted as an edge containing only three vertices ($\rightarrow tas_5$). Overall, a GroupMe! folksonomy is formally characterized via Definition 3 (cf. [1]).

Definition 3. A GroupMe! folksonomy is a 5-tuple $\mathbb{F} := (U, T, \check{R}, G, \check{Y})$, where:

- U, T, R, G are finite sets that contain instances of users, tags, resources, and groups, respectively.
- $\check{R} = R \cup G$ is the union of the set of resources and the set of groups.
- \check{Y} defines a *GroupMe! tag assignment*: $\check{Y} \subseteq U \times T \times \check{R} \times (G \cup \{\varepsilon\})$, where ε is a reserved symbol for the *empty group context*, i.e., a group that is not contained in another group when it gets tagged by a user.

In comparison to traditional folksonomies (see Definition 1), in which relations between tags mainly rely on their co-occurrences (i.e., two tags are assigned to the same resource), a GroupMe! folksonomy gains new relations between tags:

1. A relation between tags assigned from (possibly) different users to different resources, where the resources are contained in the same group.

2. A relation between tags assigned to a group g and tags assigned to resources that are contained in g .

Relations between resources become more explicit in GroupMe! folksonomies, than in traditional folksonomies. The latter allow to derive such relations if, e.g., the same tag was assigned to different resources, or if the same user has annotated different resources. In the GroupMe! system users create groups with respect to a specific topic. All resources that are arranged together in a certain GroupMe! group are related to the group's topic and are also related to each other.

The relationship between a group and the resources contained in the group can be interpreted as a *part-of-relation*. In case of constructing hierarchies among groups (groups that contain groups) further types of relations are implicated, e.g., tags that are assigned in superior might be understood as broader concepts (cf. *skos:broader* – SKOS [11]), or resources in inferior groups may be more specific than those of superior groups.

In the following section we present ranking algorithms, which exploit some of these new relations.

2.4 Ranking Strategies

In this section we present GroupMe! ranking strategies. All strategies are based on the FolkRank algorithm [2] and differ in the way GroupMe! tag assignments (which form a *4-uniform hypergraph*, cf. Definition 3) are exploited in the graph construction process. Figure 2.9 shows a tagging scenario and the hypergraph formed by the tag assignments of the scenario. The challenge of adapting the FolkRank algorithm to GroupMe! folksonomies is to identify semantically appropriate strategies for constructing a graph (*folksonomy graph*), whose adjacency matrix serves as input for the PageRank-based FolkRank algorithm.

2.4.1 FolkRank Algorithm

The core idea of the FolkRank algorithm is to transform the hypergraph formed by the traditional tag assignments (see Definition 1) into an undirected, weighted tripartite graph $\mathbb{G}_{\mathbb{F}} = (V_{\mathbb{F}}, E_{\mathbb{F}})$, which serves as input for an adaption of PageRank [12]. At this, the set of nodes is $V_{\mathbb{F}} = U \cup T \cup R$ and the set of edges is given via $E_{\mathbb{F}} = \{\{u, t\}, \{t, r\}, \{u, r\} \mid (u, t, r) \in Y\}$ (cf. Definition 1). The weight w of each edge is determined according to its frequency within the set of tag assignments, i.e., $w(u, t) = |\{r \in R : (u, t, r) \in Y\}|$ is the number of resources the user u tagged with keyword t . Accordingly, $w(t, r)$ counts the number of users who annotated resource r with tag t , and $w(u, r)$ determines the number of tags a user u assigned to a resource r . With $\mathbb{G}_{\mathbb{F}}$ represented by the real matrix A , which is obtained from the adjacency

matrix by normalizing each row to have 1-norm equal to 1, and starting with any vector \vec{w} of nonnegative reals, adapted PageRank iterates as follows:

$$\vec{w} \leftarrow dA\vec{w} + (1-d)\vec{p}. \quad (2.1)$$

Adapted PageRank utilizes vector \vec{p} as a preference vector, fulfilling the condition $\|\vec{w}\|_1 = \|\vec{p}\|_1$. Its influence can be adjusted by $d \in [0, 1]$. Based on this, FolkRank is defined as follows [2]:

Definition 4. The *FolkRank algorithm* computes a topic-specific ranking in folksonomies by executing the following steps:

1. \vec{p} specifies the preference in a topic (e.g., preference for a given tag).
2. \vec{w}_0 is the result of applying the adapted PageRank with $d = 1$.
3. \vec{w}_1 is the result of applying the adapted PageRank with some $d < 1$.
4. $\vec{w} = \vec{w}_1 - \vec{w}_0$ is the final weight vector. $\vec{w}[x]$ denotes the *FolkRank* of $x \in V$.

2.4.2 Group-Aware Ranking Strategies

To adapt the FolkRank algorithm to GroupMe! folksonomies we confine ourself on adapting the process of constructing the folksonomy graph $\mathbb{G}_{\mathbb{F}}$ from the hypergraph formed by the GroupMe! tag assignments. Therefore, we introduce three main strategies:

A. Traditional Tag Assignments. This approach reduces GroupMe! tag assignments to traditional tag assignments, as illustrated in Figure 2.10. Groups are just taken into account as resources that might or might not be tagged. Building the tripartite graph $\mathbb{G}_{\mathbb{A}} = (V_{\mathbb{A}}, E_{\mathbb{A}})$ is done analogously to FolkRank. The set of nodes and edges is given as follows: $V_{\mathbb{A}} = U \cup T \cup \check{R}$ and $E_{\mathbb{A}} = \{\{u, t\}, \{t, r\}, \{u, r\} \mid u \in U, t \in T, r \in \check{R}, g \in G \cup \{\varepsilon\}, (u, t, r, g) \in \check{Y}\}$. Computing the weight of each edge also corresponds to the FolkRank approach, e.g.: $w(u, t) = |\{(u, t, r, g) \in \check{Y} : r \in \check{R}, g \in G \cup \{\varepsilon\}\}|$ is the number of resources the user u tagged with keyword t in any group g . This strategy corresponds to the *normal FolkRank* algorithm. It just requires the preprocessing step, in which the GroupMe! folksonomy is transformed into a traditional folksonomy. The motivation of this strategy is to have a

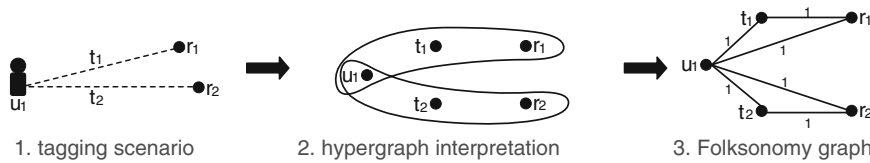


Fig. 2.10 Graph construction in traditional folksonomies. The tagging scenario represents an extract of the scenario illustrated in Fig. 2.9. The group context is ignored

benchmark strategy, which we use in order to analyze if the new group structure in folksonomies has an impact on the quality of the FolkRank algorithm.

- B. Groups as Tags.** GroupMe! users create groups about a certain topic. In general they only arrange those resources together in a group which are related to the group's topic. Resources within a same group are thus closely related to each other. The strategy "Groups as Tags" tries to emphasize this relation and creates artificial tags $t_g \in T_G$, $T_G \cap T = \emptyset$, for each group g and assigns such tags to all resources contained in g , whereby the user who added a resource to the group is declared as the *tagger*. The set of nodes is extended by T_G : $V_{\mathbb{B}} = V_{\mathbb{A}} \cup T_G$.

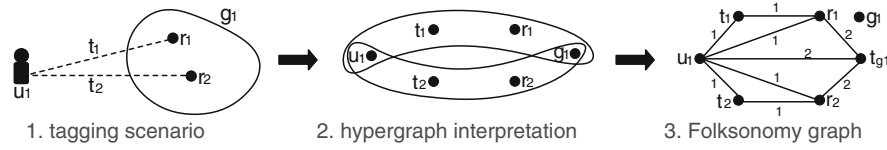


Fig. 2.11 Graph construction when interpreting groups as tags. Again, the tagging scenario represents an extract of the scenario illustrated in Fig. 2.9

The edges added to $V_{\mathbb{F}}$ by the strategy are $E_{\mathbb{B}} = E_{\mathbb{A}} \cup \{\{u, t_g\}, \{t_g, r\}, \{u, r\} | u \in U, t_g \in T_G, r \in \tilde{R}, u \text{ has added } r \text{ to group } g\}$. We use a constant value w_c to weight these edges because a resource is usually added only once to a certain group. Hence, counting, e.g., the number of users who added a resource to a specific group would not make sense.¹⁴ A hypergraph, which functions as database for this graph construction strategy, is depicted in Fig. 2.11. Here, for the group g_1 , which is treated as normal resource, a new artificial tag t_{g_1} is introduced and assigned to those resources that are member of the corresponding group.

- C. Group Context-based Tags.** The actual meaning of (possibly ambiguous) tags is hard to infer in traditional folksonomies, i.e., it is difficult to detect that a tag has ambiguous meaning, because the context of tags is only established via the users and the resources. In order to detect fuzzy usage of a tag t other tags of the users, who assigned t to a resource, and other tags, which have been assigned to resources that are tagged with t , can be utilized. The group context provides more explicit alternatives to overcome this problem. If users assign a certain tag to resources in different groups then the meaning of the tag may differ. As denoted at the end of Sect. 2.3, we can compute the degree of overlap between groups, i.e., a tag t that occurs in two groups, which do not have any other common tags and do not contain same resources, has in all probability a (slightly) different meaning depending on the group.

This strategy embeds the group context directly into the tags and replaces every tag t with a tag t_{tg} , which indicates that tag t was used in group g . It then transforms all GroupMe! tag assignments into normal tag assignment triples. For example, the GroupMe! tag assignment (u_1, t_2, r_2, g_1) , presented in Fig. 2.12, is interpreted as (u_1, t_{2g_1}, r_2) ($= tas_1$), and (u_1, t_2, r_2, g_2) is converted

¹⁴ Instead we select, e.g., $w_c(t_g, r) \approx \max(|\{u \in U : (u, t, r, g) \in \tilde{Y}\}|)$ as we believe that grouping a resource is in general more valuable than tagging it.

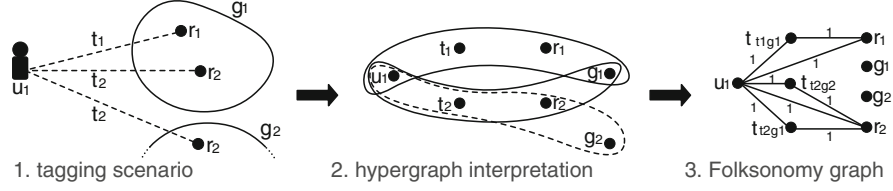


Fig. 2.12 Graph construction process when applying strategy “Group Context-based Tags” by example of an extract of the scenario illustrated in Fig. 2.9

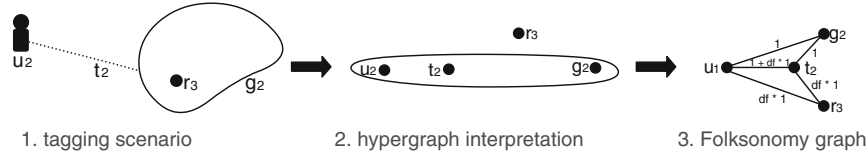


Fig. 2.13 Scenario illustrating that *group tags* are propagated attenuated by the dampen factor df

into $(u_2, t_{t_2g_2}, r_2)$ ($= tas_2$). The construction of $\mathbb{G}_{\mathbb{C}}$ is done as in the normal FolkRank algorithm, described in Sect. 2.4.1. Detecting equality of tags is the only exception, e.g., given tas_1 and tas_2 from above, the weight $w(u_1, t_{t_2g_1})$ is not only determined by tas_1 but also partially by tas_2 , although the tag $t_{t_2g_1}$ in tas_1 is not exactly equal to $t_{t_2g_2}$ in tas_2 . We compute the similarity between two tags $t_{t_xg_y}$ and $t_{t_vg_w}$ and therewith the influence of a tag assignment to a weight as follows:

\wedge	$t_x = t_v$	$t_x \neq t_v$
$g_y = g_w$	1.0	0.2
$g_y \neq g_w$	0.4	0

Hence, based on tas_1 and tas_2 it is $w(u_1, t_{t_2g_1}) = 1.4$.

In addition to the three strategies that can be applied to generate the folksonomy graph \mathbb{G} , which serves as input for the FolkRank algorithm, we present two further strategies to exploit a GroupMe! folksonomy. They can be applied as extensions to the strategies above. The motivation of both strategies is that a tag t , which was assigned to resource r or group g in certain context, is to some extent relevant to resources and groups that occur in the same context. Hence, the core idea of both strategies is to propagate tags assigned to one resource/group to other resources or groups. Such techniques synthetically increase the amount of input data and do not require to change the strategies described above substantially.

Propagation of Group Tags. GroupMe! users tag groups about 1.75 times more often than common resources [13] (see Sect. 2.2.2). By propagating tags, which have been assigned to a group (*group tags*), to the group’s resources we try to counteract this situation. For example in Fig. 2.9, tag t_2 , which is assigned to group g_2 , can be propagated to all resources contained in g_2 . An obvious benefit of this procedure is that untagged resources like r_3 obtain tag assignments (here: (u_2, t_2, r_3, g_2)). To adjust the influence of inherited tag assignments, we weight

these assignments by a dampen factor $df \in [0, 1]$. Figure 2.13 demonstrates how the folksonomy graph is constructed when group tags are propagated.

Propagation of All Tags. In the same way tags can be propagated among resources that are contained in the same group. This strategy induces propagation of (1) group tags to resources within the group, (2) resource tags of one resource to other resources within a group, and (3) resource tags to the group itself. Note that only tag assignments that have been carried out within the context of the corresponding group are considered for propagation.

2.4.3 Evaluation

Different extensions of the FolkRank algorithm have been developed and were described in the last section. To decide whether the additional information in the GroupMe! system (*the group context*), can help to improve the search performance, we have to evaluate if any of the group-aware ranking strategies, which adapt FolkRank by exploiting the group context can outperform the classical FolkRank algorithm.

2.4.3.1 Metrics

The adapted FolkRank algorithms described in Sect. 2.4.2 compute rankings for all entities of a folksonomy (users, tags, resources, and groups). In the evaluation we concentrate on ranking of resources and groups as search for resources is the most common use case of ranking in folksonomy systems. To measure the quality of our ranking strategies we used the *OSim* and *KSim* metrics as proposed in [14]. *OSim*(τ_1, τ_2) enables us to determine the overlap between the top k resources of two rankings, τ_1 and τ_2 .

$$OSim(\tau_1, \tau_2) = \frac{|R_1 \cap R_2|}{k}, \quad (2.2)$$

where $R_1, R_2 \subset \check{R}$ are the sets of resources that are contained in the top k of ranking τ_1 and τ_2 respectively, and $|R_1| = |R_2| = k$.

KSim(τ_1, τ_2), which is based on Kendall's τ distance measure, indicates the degree of pairwise distinct resources, r_u and r_v , within the top k that have the same relative order in both rankings.

$$KSim(\tau_1, \tau_2) = \frac{|\{(u, v) : \tau_1, \tau_2 \text{ agree on order of } (u, v), u \neq v\}|}{|U| \cdot (|U| - 1)}. \quad (2.3)$$

U is the union of resources of both top k rankings. τ'_1 corresponds to ranking τ_1 extended with resources R'_1 that are contained in the top k of τ_2 and not contained in τ_1 . We do not make any statements about the order of resources $r \in R'_1$ within ranking τ'_1 . τ'_2 is constructed correspondingly.

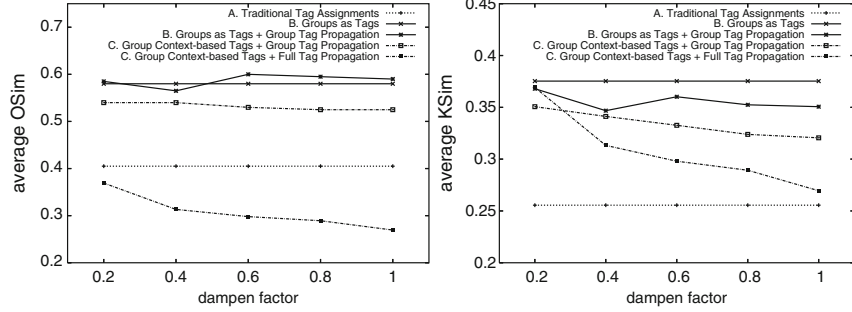


Fig. 2.14 Average OSim and KSim (with respect to 10 different top 20 ranking comparisons) for varying dampen factors, which control the influence of propagated tags, and different ranking strategies

Together, *OSim* and *KSim* are suited to measure the quality of a ranking with respect to an optimal (possibly hand-selected) ranking. Our evaluations are based on 50 hand-selected rankings: Given 10 keywords, which were out of T , and the entire GroupMe! data set, 5 experts independently created rankings for each of the keywords, which represented from their perspective the most precise top 20 rankings. By building for each keyword the average of these rankings, we gained ten optimal rankings. Among the ten keywords, there were frequently used tags as well as seldomly used ones.

2.4.3.2 Measurements and Discussion

Table 2.2 Overview of OSim and KSim for different ranking strategies ordered by OSim, where the dampen factor for propagating tags is 0.2. A denotes the “Traditional Tag Assignments” strategy, B is the “Groups as Tags” strategy, and C is the “Group Context-Based Tags”

	Strategy	OSim	KSim
(i)	C + Full Tag Propagation	0.610	0.369
(ii)	B + Group Tag Propagation	0.585	0.368
(iii)	B	0.580	0.375
(iv)	C + Group Tag Propagation	0.540	0.351
(v)	B + Full Tag Propagation	0.465	0.273
(vi)	A	0.405	0.255
(vii)	C	0.390	0.257
(viii)	A + Group Tag Propagation	0.360	0.237
(ix)	A + Full Tag Propagation	0.345	0.247

Table 2.2 gives an overview on the measured results for each ranking strategy introduced in Sect. 2.4.2 with respect to OSim and KSim metrics. The strategies are ordered according to their OSim performance, whereas both OSim and KSim values

are averaged out of ten test series (for the ten different keywords and corresponding hand-selected rankings). In terms of the OSim, “C – Group Context-Based Tags” can be identified as best strategy: It computes rankings, which contain 61% of the resources that also occur in the corresponding hand-selected top 20 ranking lists. Group tag propagation does not influence the approach “B – Groups as Tags” strongly as strategies (ii) and (iii) have nearly the same OSim values. This can be explained with the functionality of “B”: For each group, “B” introduces artificial

Table 2.3 Top 10 rankings computed by different ranking strategies (and by hand, respectively) for the term “*socialpagerank*”.

Rank	Hand-selected	A. Traditional TAS	B. Groups as Tags	C + Full Tag Propagation
1.	Optimizing Web search using social annotations	Optimizing Web search using social annotations	Optimizing Web search using social annotations	Yahoo! research
2.	Exploring social annotations for the sem. . .	The Semantic Web: Will It All End In Tiers?	HITS	Optimizing web search using social annotations
3.	Personalized PageRank	New *Semantic* Web!	Webpage Ranking (<i>group</i>)	Ontologies are us
4.	SimRank	The Semantic Web: An Introduction	SimRank	Bibsonomy
5.	PageRank	The Semantic Web: Scientific American	PageRank	HITS
6.	FolkRank	LEGOLAND	Topic-sensitive PageRank	SimRank
7.	Ontologies are us	eschbach (<i>group</i>)	Personalized PageRank	PageRank
8.	Topic-sensitive PageRank	Andreas Eschbach Wikipedia	Yahoo! research	Topic-sensitive PageRank
9.	Bibsonomy	Andreas Eschbach Homepage	FRank	Personalized PageRank
10.	FRank	Andreas Eschbach: Der Nobelpreis	Ontologies are us	FolkRank

tags and assigns those tags to the group’s members. Considering the graph structure, this almost conforms to propagating the tags of a group to its members.

Strategy (vi) does not exploit the group structure as it reduces GroupMe! tag assignments to traditional tag assignments (see Sect. 2.4.1) and can therewith be interpreted as the traditional FolkRank algorithm. The extensions of FolkRank, (viii) and (ix), which rudimentary exploit the group structure, do not improve the overlapping similarity of 0.405 but rather degrade the quality of FolkRank. We assume that the approach of propagating tags without modeling the group dimension within the graph, which serves as input for the ranking algorithm, primarily increases the recall but has a negative effect on the precision.

Regarding the KSim, strategy (iii), which treats groups as tags, performs best, followed by strategies (i), (ii), and (iv). The quality of the strategies (i)–(iv) is, in view of KSim, more than 30% better than the quality of strategies (v)–(ix).

Figure 2.14 gives an idea about how the ranking strategies behave when varying the dampen factor for tag propagation. Naturally, the dampen factor does not effect strategies “A – Traditional Tag Assignments” and “B – Groups as Tags” because both strategies do not make use of tag propagation. When varying the dampen factor, the OSim value is comparatively constant as well as for the strategies that base on propagation of group tags. The OSim and KSim of strategy “C + Full Tag Propagation” continuously degrades when the dampen factor increases. Gazing at the idea of “Full Tag Propagation” illustrates this behavior: Assume there is a resource r in a group g , which contains 20 other resources, and r is the only resource, which is tagged with t . Then, propagation of t to g and all resources of g with a dampen factor of 1.0 would conceal the prominent role of resource r in terms of tag t . Hence, ranking the resources of g in an adequate order gets difficult (see KSim value), and the increased recall complicates the process of identifying resources to put into the top k of the ranking for tag t .

Table 2.3 outlines example rankings computed for the tag “socialpagerank” by different ranking strategies. Furthermore, it lists the corresponding hand-selected ranking, which is based on votings of five experts. Within the GroupMe! data set the resource “*Optimizing web search using social annotations*,” a paper which proposes the SocialPageRank algorithm, was the only resource tagged with the keyword “socialpagerank.” This resource was ranked at first place in the hand-selected ranking, and almost every ranking strategy conforms to this decision. Starting from the second position the ranking of strategy “A,” which represents the traditional FolkRank algorithm, gets imprecise. As strategy “A – Traditional Tag Assignments” does not exploit the group structure, the only solution to discover other relevant resources rests upon the users, who annotated the resource, and other tags that have been assigned to the resource. The group-based ranking strategies, on the other hand, are able to detect adequate resources via the group containing the resource. In the given example, this group is “*Webpage Ranking*” and strategy “B – Groups as Tags” is the only strategy that lists the group also within the top ten.

2.4.3.3 Results

The goal of our investigation was to identify whether grouping of resources in folksonomies has an impact on the quality of search strategies in social networks. To give proof on our hypothesis that grouping improves the quality of search, it is necessary to compare the search strategies that explore the grouping context to those search strategies which do not. As benchmark, we have chosen the FolkRank algorithm and have developed search algorithms that extend FolkRank to exploit the group context as described in Sect. 2.4.1. All algorithms, FolkRank as well as the group-aware extensions, were tested under the same conditions, i.e., the same set of data, hardware, etc.

We tested our hypothesis with a one-tailed t -test. The null hypothesis H_0 is that some group-aware FolkRank extension is as good as a the normal FolkRank without

group- awareness, and tested it with a significance level of $\alpha = 0.05$. Tests were performed for the two measures OSim and KSim (see Sect. 2.4.3):

OSim. With respect to OSim, the strategy “B – Groups as Tags” is significantly better than normal FolkRank.

Furthermore, FolkRank did not improve if we applied any of the tag-propagation strategies described in Sect. 2.4.2, and, indeed, the strategy “B – Groups as Tags” was significantly better than normal FolkRank with or without tag propagation. The variations of “B – Groups as Tags” to reflect tag propagation were, one compared to the other, not significantly different, but only the propagation of group tags showed significant improvement in comparison to FolkRank (with or without tag propagation).

Also the strategy “C – Group Context-based Tags,” where full propagation of tags was performed (damping factor 0.2), was significantly better than the normal FolkRank regardless of whether any propagation of tags was performed in the latter. From our test data, we hypothesize that strategy “C” benefits from the propagation of tags while “B” does not. Our actual data did not give statistically significant results on this, and we will investigate the impact on tag propagation in our future work.

KSim. With respect to KSim, the strategy “B – Groups as Tags” is significantly better than normal FolkRank, whether or not the latter uses any tag propagation strategy.

Also the strategy “C – Group Context-Based Tags,” where group tags are propagated (damping factor 0.2) is significantly better than normal FolkRank, whether or not the latter uses any tag propagation strategy.

OSim and KSim. Only the strategy “B – Groups as Tags” (without tag propagation or with group tag propagation, damping factor 0.2) was significantly better with respect to both measures, OSim and KSim, than normal FolkRank (whether or not the latter uses any tag propagation strategy).

Our evaluation indicated that the grouping of resources significantly improves the quality of search in folksonomies. The grouping activity itself brings many advantages for users: they can organize resources of interest, they can overlook and inspect a group’s content, they can share groups with fellow users, and can explore the information in a folksonomy in novel ways, e.g., by requesting new, artificial groups that collect contents of all groups for the same topic, that collect the most popular groups or resource, etc. Furthermore, the drag & drop metaphor realized in the GroupMe! system makes the grouping activity intuitive for users, and from our experience with running GroupMe! we have seen that users like grouping [13]. Thus, while grouping is an easy and well-received feature for folksonomies, this activity provides, on the technical side, valuable information to detect relevant resources, and to improve the quality of search, and seems to be a very promising approach to improve current folksonomy systems.

2.5 Related Work

With the advent of Web 2.0 and its new design patterns, which are proposed in [15], social tagging systems like del.icio.us,¹⁵ Flickr,¹⁶ or Last.fm¹⁷ became quite successful. In [4], the authors developed a *tagging system design taxonomy*, which allows to characterize such tagging systems regarding different dimensions. Table 2.4 summarizes characteristics of GroupMe! according to this design taxonomy and compares them with three related tagging systems: (1) BibSonomy [9] is a social

Table 2.4 GroupMe! tagging design in comparison to other social tagging systems. And user incentives for tagging

Dimension/system	<i>GroupMe!</i>	<i>BibSonomy</i>	<i>del.icio.us</i>	<i>Flickr</i>
<i>Tagging rights</i>	Free-for-all	Free-for-all	Free-for-all	Permission-based
<i>Tagging support</i>	Blind/viewable	Suggested	Suggested	viewable
<i>Aggregation model</i>	Bag	Bag	Bag	Set
<i>Object type</i>	Multimedia	Textual	Textual	Images
<i>Source of material</i>	Global	Global	Global	User-contributed
<i>Social connectivity</i>	Links	Links, Groups	Links	Links
<i>Resource connectivity</i>	Groups	None	None	Groups
User incentives	future retrieval contribution sharing attract attention self presentation	future retrieval contribution sharing	future retrieval contribution sharing attract attention	future retrieval contribution sharing attract attention self presentation

bookmarking and publication sharing platform, (2) del.icio.us is currently the most used social bookmarking platform, and (3) Flickr is a platform, which enables users to upload and share photos. GroupMe! system characteristics and differences to the three other systems are described as follows.

Tagging rights. GroupMe! allows every user to tag everything (*free-for-all*) as this enables us to gather more tags about a resource and also a higher variety of keywords than in constrained systems such as Flickr, which restricts tagging, e.g., to the resource owner, or her friends.

Tagging support. In the GroupMe! system users always annotate resources within the context of a group. During the tagging process they are not supported with tag suggestions. However, users have the ability to gaze at tags that have already been assigned to resources in context of the actual group. Tags that have been assigned in context of other groups are not visible to the user when tagging as those tags are possibly not adequate to the actual group context. Regarding tagging support GroupMe! allows for both, *blind* and *viewable* tagging. BibSonomy, del.icio.us, and other systems provide tag suggestions to the users, which makes tagging easier for users, but limits, in our opinion, the variety of tags that are assigned to resources.

¹⁵ <http://del.icio.us>

¹⁶ <http://flickr.com>

¹⁷ <http://last.fm>

Aggregation model. In comparison to Flickr, which does not allow for duplicated tags (*set*), GroupMe! allows different users to assign the same tag to a certain resource (*bag*). The aggregation model has a strong impact on the structure of the evolving folksonomy [5], which is, in simple terms, a collection of tag assignments. In [6], Vander Wal differentiates between *narrow* folksonomies, which evolve in tagging systems such as Flickr, and *broad* folksonomies, which arise in systems such as del.icio.us or GroupMe!. Moreover, the structure of a folksonomy influences the choice of appropriate search and ranking strategies. In this article we presented ranking strategies that are optimized for broad folksonomies.

Object type. GroupMe! is the only system listed in Table 2.4 that supports tagging of resources displayed in a multimedia fashion. Although systems such as del.icio.us enable users to bookmark and tag arbitrary Web resources, they just visualize resources in a textual way. Hence, while tagging, e.g., a video in del.icio.us, users are not able to watch the video they tag within the del.icio.us system, but have to visit the corresponding Web site. *CombinFormation* [16] – a system which also allows (re)organizing Web content – provides similar functionality regarding visualization of resources, however, neither provides tagging functionality nor makes use of the new structures to provide enhanced search and browsing functionalities.

Source of material. Resources that can be annotated and grouped in GroupMe! are globally distributed over the Web and referenced by their URL. This enables GroupMe! to handle often-changing resources such as RSS feeds appropriately: Whenever a group is accessed, the most recent versions of the contained resources are displayed. From the perspective of *source of material* GroupMe! rather adheres to the idea of social bookmarking than to systems such as Flickr or YouTube, which enable users to upload and publish own content.

Social connectivity. All systems listed in Table 2.4 allow users to be linked together. GroupMe! does not provide integrated features, but utilizes users' FOAF descriptions in order to identify links between users.

Resource connectivity. Independent of the users' tags, a few resource sharing systems provide other features to connect resources. There are some systems that allow users to organize themselves into groups and that provide functionality to retrieve resources, which are related to these groups – e.g., BibSonomy or Connotea.¹⁸ Furthermore, del.icio.us allows users to connect tags by building so-called *tag bundles*. However, to the best of our knowledge, Flickr and GroupMe! are at the moment the only notable tagging systems that enable users to assign resources to groups explicitly. GroupMe! groups differ from Flickr groups in two relevant aspects: (1) Flickr groups are simple sets of images unlike GroupMe! groups, which capture arbitrary Web resources of interest and can be fashioned by the users (resources can be resized and arranged), and (2) GroupMe! supports, in contrast to Flickr, tagging of groups.

¹⁸ <http://www.connotea.org>

User incentives. GroupMe! users have several motivations to annotate resource ranging from simplification of future retrieval to self-presentation (e.g., some users tag resources with *holiday* in order to express which locations they have visited).

Folksonomies represent the database of tagging systems. They evolve, according to [17], like *desire lines* over time. Visualizing such temporal formation is discussed in [18] and demonstrated with *Yahoo! TagLines*.¹⁹ A basic formal folksonomy model is presented in [7, 8, 19]. In [10], Wu et al. extend this model with a time dimension. The GroupMe! folksonomy extends the folksonomy model defined in [9] by adding a group context to tag assignments (see Sect. 2.3). Therewith, new relations between resources, groups, and tags emerge that can be exploited by search and ranking algorithms as we show in this article. Search and ranking algorithms that operate on traditional folksonomies have already been successfully applied in order to improve Web search. In [20], the authors introduce *SocialSimRank*, which adapts *SimRank* [21] and computes similarity between tags and resources, respectively. Furthermore, Bao et al. presented the *SocialPageRank* algorithm, which ranks Web resources according to how popular they are annotated. *FolkRank* [2] is another folksonomy-based search algorithm, which adapts the well-known *PageRank* [12] algorithm and involves user preferences. In this article we described how FolkRank can be applied to GroupMe! folksonomies in order to improve the quality of rankings.

Learning relations between tags is another challenge in social tagging systems that can be utilized to improve retrieval of resources additionally. Hotho et al. presented an approach to mine association rules in folksonomies that point to *subtag-supertag* relations [22]. The GroupMe! folksonomy model provides a foundation to deduce such relations more precisely, e.g., by analyzing tags that have been assigned to a group and tags assigned to the group members. In [23], the authors investigated how to learn more concrete semantics from folksonomies. In particular, they presented an approach to distinguish between *event tags* and *place tags*. Such approaches for learning semantics can also be applied to GroupMe!.

In addition to analysis of emerging semantics, GroupMe! also focuses on explicit combination of Web 2.0 and Semantic Web technologies. Instead of proposing that the Semantic Web, as envisioned by Berners-Lee et al. in [24] and specified by the W3C Semantic Web Activity,²⁰ is *dead* as provocatively stated by Naaman during a panel discussion at WWW '07 [25]) we follow [26] and believe that bringing both technologies together will originate the future Web. Therefore, the GroupMe! system is implemented as an RDF aggregator and generator, and provides, in addition to RDF and RSS feeds, RESTful interfaces and corresponding client APIs to access these RDF data. In this way, GroupMe! conforms to the *Linked Data* principles outlined in [27]. Most of the other systems such as *CiteULike*²¹ or *BibSonomy* just offer RSS export, or deliver data via APIs using application-specific vocabularies to

¹⁹ <http://research.yahoo.com/taglines>

²⁰ <http://www.w3.org/2001/sw/>

²¹ <http://www.citeulike.org>

describe data. For example, Flickr provides interfaces to access their data corpus, which return XML- or JSON-formatted data using a Flickr-specific vocabulary instead of referring to well-defined ontologies such as *Friend-of-a-friend* vocabulary [28], e.g., they use “photo” instead of “foaf:Image”.

Semantic Wikis [29] and Semantic Blogs [30] prevent lack of semantically well-defined content as they oblige users to link content with ontology concepts. With GroupMe! we do not want to burden the user with knowledge engineering tasks and thus do not foresee such functionality at the moment, but rather plan to make use of ontology learning strategies as proposed in [31]. In [10], Wu et al. present a probabilistic approach to derive semantics from tags assignments, i.e., they determine the conditional probability that a tag refers to a concept (*conceptual space*), where the user, who utilized a tag, represents the pre-condition.

Tagging systems furthermore provide a convenient base for user modeling and personalization functionalities by utilizing tag-based user profiles. In [32], the authors propose an algorithm to learn such user profiles and in [33] they show how to adapt tag-based user profiles over time. The benefit of tag-based user profiles is demonstrated by Firan et al. in [34], where they show that tag-based profiles outperform track-based profiles in order to recommend music tracks to a user. Similar strategies can be exploited by the GroupMe! system in order to recommend content within the GroupMe! system. In this article we present ranking strategies, which can be applied in order to personalize content.

2.6 Conclusions and Future Work

We presented GroupMe!, a social bookmarking and tagging system combining Semantic Web and Web 2.0 techniques. We outlined the innovative character of GroupMe!, characterized by a novel drag & drop based user interface, support for arbitrary multimedia resources, and the feature of grouping resources. We have shown how Web 2.0 systems such as GroupMe! can utilize different data sources (by means of RDF Metadata extractors or RDF search engines) to enrich the resources of the system’s data corpus with additional RDF descriptions. We also described our API, which enables other applications to make use of the GroupMe! data. Regarding system usage, our evaluations approved that users appreciate both, the grouping functionality and the comfortable integration of multimedia resources.

We extended the classical folksonomy, containing *users*, *resources*, and *tags* by the *group* context. Based on this extended folksonomy, we proposed different strategies as to how the folksonomy ranking algorithm FolkRank can be extended to take the group context into account in order to improve the search performance in group-aware folksonomies. Our evaluations have shown that the ranking algorithms taking the group context into account perform significantly better than the classical FolkRank algorithm.

In the future, we plan to exploit the group structure of the GroupMe! system in different directions: First, we want to simplify the creation process of a group.

Therefore, we engage link prediction algorithms and recommender systems to automatically group resources that are relevant to a user and are related to a specific topic according to their former group membership. Afterwards users can extend and/or modify this group, which can be interpreted as feedback regarding the quality of the recommendations.

Second, the visualization capabilities of GroupMe! will be extended. We consider on one hand *zoomable interfaces*²² where a large content can be visualized in different degrees of detail. On a global view, clusters of groups with similar topics can be displayed, enabling users to zoom into a more detailed level visualizing groups. When a user zooms into the group the contained resources and groups become visible. Such a zoomable interface enables users to see the content of the whole GroupMe! system at a glance while the content of any resource is accessible in a few zoom operations. On the other hand, we investigate automatic arrangement techniques for groups. For example, algorithms can be implemented that take usage statistics into account to emphasize important resources in a group by resizing them or rearranging them.

The third direction of extending the GroupMe! system aims on embedding the GroupMe! system into the Web 2.0 sphere. While first steps have already been taken by automatically searching and extracting RDF metadata for GroupMe! resources and providing an API that gives the GroupMe! data back to the Web 2.0 community, we think of techniques that actively push data into the Web 2.0 sphere. Therefore, we plan to improve the integration of services such as del.icio.us, Flickr, CiteULike, or Bibsonomy. For example, when annotating Web resources within the GroupMe! system, we want to give users, who have a del.icio.us account, the opportunity to decide whether their tag assignments should also be propagated to del.icio.us, and vice versa. To set a good role model, we currently extend the GroupMe! system with additional semantically described interfaces, which allow for both, querying and adding/updating GroupMe! data. The RDF (Meta) search engine, described in Sect. 2.2.1, is, for instance, accessible as Semantic Web Service by making use of OWL-S and REST principles.

Enhancing Web 2.0 systems with Semantic Web technologies is, in our opinion, an adequate strategy to realize the visions associated with the Semantic Web, e.g., most Web 2.0 systems make their services available via API so that describing these interfaces semantically would be feasible. GroupMe! demonstrates the benefits of combining Web 2.0 and Semantic Web technologies. GroupMe! brings Web 2.0 and Semantic Web technologies together and reveals the benefits of combining both techniques. It aggregates semantic descriptions about resources, captures the semantics of user interactions, and illustrates how semantic relations between resources, gained by the group context, improve search and ranking strategies.

Acknowledgements We thank Nicole Ullmann, Mischa Frank, Daniel Plappert, Patrick Siehndel, and Zhivko Asenov for their contribution and engagement in realizing the GroupMe! system.

²² <http://www.zoomorama.com>

References

1. Abel, F., Frank, M., Henze, N., Krause, D., Plappert, D., Siehndel, P.: GroupMe! – Where Semantic Web meets Web 2.0. In: Int. Semantic Web Conference (ISWC 2007) (November 2007)
2. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: FolkRank: A ranking algorithm for folksonomies. In: Proc. of Workshop on Information Retrieval 2006 of the Special Interest Group Information Retrieval (FGIR 2006), Hildesheim, Germany (October 2006)
3. Fielding, R.T., Taylor, R.N.: Principled design of the modern web architecture. In: Proc. of the 22nd Int. Conf. on Software Engineering (ICSE '00), New York, NY, USA, ACM Press (2000) 407–416
4. Marlow, C., Naaman, M., Boyd, D., Davis, M.: HT06, tagging paper, taxonomy, flickr, academic article, to read. In: Proc. of the 17th Conf. on Hypertext and Hypermedia (HYPERTEXT '06), New York, NY, USA, ACM Press (2006) 31–40
5. Vander Wal, T.: Folksonomy. <http://vanderwal.net/folksonomy.html> (July 2007)
6. Vander Wal, T.: Explaining and showing broad and narrow folksonomies. http://www.personalinfocloud.com/2005/02/explaining_and_.html (February 2005)
7. Halpin, H., Robu, V., Shepherd, H.: The complex dynamics of collaborative tagging. In: Proc. of 16th Int. World Wide Web Conference (WWW '07), New York, NY, USA, ACM Press (2007) 211–220
8. Mika, P.: Ontologies are us: A unified model of social networks and semantics. In: Proc. Int. Semantic Web Conference (ISWC 2005). (November 2005) 522–536
9. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: BibSonomy: A social bookmark and publication sharing system. In: de Moor, A., Polovina, S., Delugach, H., eds.: Proc. First Conceptual Structures Tool Interoperability Workshop, Aalborg (2006) 87–102
10. Wu, X., Zhang, L., Yu, Y.: Exploring social annotations for the Semantic Web. In: Proc. of 15th Int. World Wide Web Conference (WWW '06), New York, NY, USA, ACM Press (2006) 417–426
11. Brickley, D., Miles, A.: SKOS Core Vocabulary Specification. W3C working draft, W3C (November 2005) <http://www.w3.org/TR/swbp-skos-core-spec>
12. Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank citation ranking: Bringing Order to the Web. Technical report, Stanford Digital Library Technologies Project (1998)
13. Abel, F., Henze, N., Krause, D.: A Novel Approach to Social Tagging: GroupMe! In: 4th Int. Conf. on Web Information Systems and Technologies (WEBIST). (May 2008)
14. Haveliwala, T.H.: Topic-sensitive PageRank: A context-sensitive ranking algorithm for Web search. *IEEE Transactions on Knowledge and Data Engineering* **15**(4) (2003) 784–796
15. O'Reilly, T.: What is web 2.0? – design patterns and business models for the next generation of software (September 2005)
16. Kerne, A., Koh, E., Dworaczyk, B., Mistrot, J.M., Choi, H., Smith, S.M., Graeber, R., Caruso, D., Webb, A., Hill, R., Albea, J.: combinFormation: A mixed-initiative system for representing collections as compositions of image and text surrogates. In: Proc. of the ACM/IEEE Joint Conference on Digital Libraries (JCDL 2006), Chapel Hill, NC, USA, ACM Press (June 2006) 11–20
17. Merholz, P.: Metadata for the masses. Adaptive Path (October 2004)
18. Dubinko, M., Kumar, R., Magnani, J., Novak, J., Raghavan, P., Tomkins, A.: Visualizing tags over time. In: Proc. of 15th Int. World Wide Web Conference (WWW '06), New York, NY, USA, ACM Press (2006) 193–202
19. Marlow, C., Naaman, M., Boyd, D., Davis, M.: Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead. In: Collaborative Web Tagging Workshop at WWW '06. (May 2006)
20. Bao, S., Xue, G., Wu, X., Yu, Y., Fei, B., Su, Z.: Optimizing Web search using social annotations. In: Proc. of 16th Int. World Wide Web Conference (WWW '07), New York, NY, USA, ACM Press (2007) 501–510
21. Jeh, G., Widom, J.: SimRank: A measure of structural-context similarity. In: Proc. of Int. Conf. on Knowledge Discovery and Data Mining (SIGKDD), Edmonton, Alberta, Canada, ACM Press (July 2002)

22. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Emergent Semantics in BibSonomy. In Hochberger, C., Liskowsky, R., eds.: *Informatik 2006: Informatik für Menschen*. Volume 94(2) of LNI., Bonn, GI (October 2006)
23. Rattenbury, T., Good, N., Naaman, M.: Towards automatic extraction of event and place semantics from flickr tags. In: *Proc. of the 30th Int. ACM SIGIR Conf. on Information Retrieval (SIRIR '07)*, New York, NY, USA, ACM Press (2007) 103–110
24. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. *Scientific American* **284**(5) (2001) 34–43
25. Naaman, M.: The Semantic Web is dead. In: *Panel Discussion: The Role of Multimedia Metadata Standards in a (Semantic) Web 3.0*, 16th Int. World Wide Web Conference (WWW '07). (May 2007)
26. Ankolekar, A., Krötzsch, M., Tran, T., Vrandečić, D.: The two cultures: Mashing up Web 2.0 and the Semantic Web. In: *Proc. of 16th Int. World Wide Web Conference (WWW '07)*, New York, NY, USA, ACM Press (2007) 825–834
27. Berners-Lee, T.: Linked Data – design issues. Technical report, W3C (May 2007) <http://www.w3.org/DesignIssues/LinkedData.html>
28. Brickley, D., Miller, L.: FOAF Vocabulary Specification 0.91. Namespace document, FOAF Project (November 2007) <http://xmlns.com/foaf/0.1/>
29. Oren, E., Völkel, M., Breslin, J.G., Decker, S.: Semantic Wikis for personal knowledge management. In: Bressan, S., Küng, J., Wagner, R., eds.: *Proc. of the 17th Int. Conf. on Database and Expert Systems Applications (DEXA 2006)*. Volume 4080 of LNCS., Kraków, Poland, Springer (September 2006) 509–518
30. Cayzer, S.: Semantic blogging and decentralized knowledge management. *Commun. ACM* **47**(12) (December 2004) 47–52
31. Cimiano, P., Pivk, A., Schmidt-Thieme, L., Staab, S.: Learning taxonomic relations from heterogeneous sources of evidence. In: *Ontology Learning from Text: Methods, Evaluation and Applications*. *Frontiers in AI*. IOS Press (2005) 59–73
32. Michlmayr, E., Cayzer, S.: Learning user profiles from tagging data and leveraging them for personal(ized) information access. In: *Proc. of the Workshop on Tagging and Metadata for Social Information Organization*, 16th Int. World Wide Web Conference (WWW '07). (May 2007)
33. Michlmayr, E., Cayzer, S., Shabajee, P.: Add-A-Tag: Learning Adaptive user profiles from bookmark collections. In: *Proc. of the 1st Int. Conf. on Weblogs and Social Media (ICWSM '06)*. (March 2007)
34. Firan, C.S., Nejdl, W., Paiu, R.: The benefit of using tag-based profiles. In: *Proc. of the 2007 Latin American Web Conference (LA-WEB 2007)*, Washington, DC, USA, IEEE Computer Society (2007) 32–41

Web 2.0 & Semantic Web

Devedžić, V.; Gašević, D. (Eds.)

2009, XX, 206 p. 52 illus., Softcover

ISBN: 978-1-4419-1218-3